Cognition 154 (2016) 95–101

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Object formation in visual working memory: Evidence from object-based attention

Jifan Zhou *, Haihang Zhang, Xiaowei Ding, Rende Shui, Mowei Shen *

Department of Psychology and Behavioral Sciences, Zhejiang University, PR China

ARTICLE INFO

Article history:
Received 25 January 2016
Revised 12 April 2016
Accepted 14 April 2016

Keywords:
Visual working memory
Visual perception
Information integration
Object-based attention

ABSTRACT

We report on how visual working memory (VWM) forms intact perceptual representations of visual objects using sub-object elements. Specifically, when objects were divided into fragments and sequentially encoded into VWM, the fragments were involuntarily integrated into objects in VWM, as evidenced by the occurrence of both positive and negative object-based attention effects: In Experiment 1, when subjects' attention was cued to a location occupied by the VWM object, the target presented at the location of that object was perceived as occurring earlier than that presented at the location of a different object. In Experiment 2, responses to a target were significantly slower when a distractor was presented at the same location as the cued object (Experiment 2). These results suggest that object fragments can be integrated into objects within VWM in a manner similar to that of visual perception.

1. Introduction

To obtain coherent and stable visual representations of the outside world, the mental system must overcome the problem of discrete inputs caused by occlusions, saccades, and motions; to that end, visual working memory (VWM) contributes greatly because of its role in the brief storage and manipulation of visual information (Baddeley & Hitch, 1974). VWM has been suggested to support the reassembling of sequentially presented slices of a figure into a complete mental image during anorthoscopic perception (Parks, 1965), and can enable persisting object representations to survive visual disruption, motion, and featural changes during the tunnel effect (Burke, 1952; Flombaum & Scholl, 2006). Thus, one important function of VWM appears to be supporting online perception by providing storage for intermediate representations produced during visual processing. This idea was explicitly proposed by Gao, Gao, Li, Sun, and Shen (2011) in their interactive model of VWM, which emphasizes VWM's role as a buffer for visual processing: namely, the intermediate representations produced by visual processing are saved into VWM for further use, such as for integration with subsequent representations or modulation according to new inputs. In this way, the final coherent representation is constructed in an "incremental" manner (Marr, 1982; Roelfsema, 2005; Ullman, 1984). The interactive model proposes that the laws of visual perception do not merely govern perception itself, but also constrain the storage mechanism of VWM, as these processes intermix during visual processing. Gao et al. (2011) provided evidence for this hypothesis by demonstrating that the outputs of parallel perception can be automatically selected into VWM and then maintained as integrated objects, but those requiring serial attentive processing cannot. Our recent study (Shen, Xu, Zhang, Shui, Zhang, & Zhou, 2015) provided direct evidence for the interactive model, showing that sequentially presented figure components of the Ponzo illusion[1] figure, when encoded in VWM, could be involuntarily integrated to form an intact figure, which in turn led to a VWM version of the Ponzo illusion. This suggests that further processing will automatically implement on the contents stored in VWM if they can be integrated to form a meaningful representation.

Given these previous findings, we focused on *visual objects* in the present study to further investigate the information integration occurred in VWM. More specifically, we examined whether visual elements stored in VWM are involuntarily integrated to form a visual object. Visual object can be defined at different levels in a hierarchical description of an image (Feldman, 2003; Scholl, 2001). In this paper, object refers to a simple type of object that has closed contour, which has been frequently used in previous studies about object-based attention (e.g., Alvarez & Scholl, 2005; Egly, Driver, & Rafal, 1994; Hollingworth, Maxcey-Richard, & Vecera, 2012; Marino & Scholl, 2005). Visual objects are considered

* Corresponding authors at: Department of Psychology and Behavioral Sciences, Xixi Campus, Zhejiang University, Hangzhou 310007, PR China.
E-mail addresses: jifanzhou@zju.edu.cn (J. Zhou), mwshen@zju.edu.cn (M. Shen).

[1] The Ponzo illusion refers to the phenomenon that perception of the length of horizontal lines is distorted when they are presented across a pair of converging lines.

http://dx.doi.org/10.1016/j.cognition.2016.04.009
0010-0277/

the fundamental units of visual processing (Chen, 2012; Scholl, 2001; Strickland & Scholl, 2015); indeed, they are proposed to be the units of VWM capacity (Lee & Chun, 2001; Luck & Vogel, 1997). Even in studies that held a different view (Alvarez & Cavanagh, 2005; Awh & Vogel, 2007; Bays & Husain, 2008, 2009)—namely, that VWM capacity should be measured by the amount of information rather than the number of objects—the materials to be memorized were still visual objects, and the amount of information was usually manipulated by varying the complexity of these objects. Thus, it seems that most modern VWM studies implicitly assume that VWM stores visual information in the form of objects. Then, is it possible that visual objects, the basic units of VWM, can be formed inside VWM? According to the interactive model (Gao et al., 2011), VWM is involved in a variety of perceptual processes. Reasonably, object processing is included; thus, VWM could theoretically store visual elements for assembling an object, and integrate them while following perceptual rules to create objects in VWM. Furthermore, since most of the visual processes occur automatically without top-down intention, if VWM serve visual processing as a buffer to save intermediate representations, as suggested by the interactive model, the VWM integration of objects should be able to occur involuntarily. Our previous work (Shen et al., 2015) demonstrated involuntary integration occurring at the level of figure (which is composed by a set of simple objects), but the conclusion cannot be directly generalized to more fundamental unit—object. Or to say, in Feldman's (2003) term for visual objects, we aimed to generalize the involuntary VWM integration from disjoint regular subtrees (generically related visual items) to lower-level objects (defined by spatial relationship of visual elements, such as closure) of the visual hierarchy of an image. The figure including disjoint regular subtrees, or the high-level objects with complex structures, are usually able to be reorganized and reinterpreted to have a new form. Such reorganization is easily affected by top-down factors (Scholl, 2001). Thus, it is possible that the involuntary VWM integration only occurs for the high-level objects (or figures) that can be affected by top-down manipulations, not for the low-level objects defined primarily by bottom-up factors (such as those defined by closed contour) since such objects could be handled by the encapsulated perceptual modules that are independent of VWM. However, whether involuntary VWM integration occurs for simple objects has not been empirically examined.

To test this hypothesis, we sequentially presented object fragments to participants, thereby creating a situation in which VWM is needed for object integration. Furthermore, we used object-based attention effects as measures of whether the objects had indeed been formed in VWM. A considerable number of studies on object-based attention have revealed that attention automatically spreads within an object, and it is difficult to focus on a single part while neglecting the rest parts of the object (Chen, 2012; Duncan, 1984; Egly et al., 1994; Eriksen & Eriksen, 1974). This feature of object-based attention can have both positive and negative effects. The positive effect is that processing of one part of an object may be facilitated when attention is cued to other parts of that same object. For example, when participants make temporal order judgments (TOJs), the target appearing inside the cued object is typically jugged as appearing earlier than the target simultaneously appearing inside the uncued object. This is believed to be due to increased transmission speed caused by object-based attention (Abrams & Law, 2000; Chen, 2012; Duncan, 1984; Eriksen & Eriksen, 1974; Shore, Spence, & Klein, 2001; Stelmach & Herdman, 1991). In contrast, the negative effect refers to that it can be hard to selectively inhibit interference from distractors when the distractors appear inside the same cued object as the target (Driver, Davis, Russell, Turatto, & Freeman, 2001). In other words, distractors will have a stronger interfering effect on target recognition (i.e., will make responses slower) when they appear inside same object than inside different object (Baylis & Driver, 1992; Chen & Cave, 2008; Harms & Bundesen, 1983; Kramer & Jacobson, 1991). Focusing on these two types of object-based attention effect, we designed experiments to examine whether object fragments stored in VWM are integrated into whole objects. Specifically, participants were required to memorize sequentially presented object fragments, then perform a TOJ (Experiment 1) or interference (Experiment 2) task to measure the positive and negative "same-object"[2] effects, respectively. If the elements were combined into objects in VWM, and therefore modulated attention, we could expect the following to occur. First, in Experiment 1, when attention is cued to the location of an object that has been stored in VWM, the target in the location of the cued object (i.e., appearing as if inside the VWM object) would be perceived as appearing earlier than the target synchronously presented in the location of an uncued object. Second, in Experiment 2, when a distractor appeared at the location of an object stored in VWM, responses to target presented inside the "same object" would be much slower due to interference from the distractor.

2. Experiment 1

2.1. Participants

Twenty participants were recruited from undergraduates of Zhejiang University. They had normal or corrected-to-normal vision, and were naïve to the aim of the study. They gave informed consent and received financial compensation for their participation. The study was approved by the institutional review board at the Department of Psychology and Behavioral Science, Zhejiang University.

We decided on the sample size before we began collecting data. Our pilot work (identical to this experiment, except that the mask after the memory items displays were removed) got a fairly large effect size (Cohen's $d = 0.59$, $N = 10$). According to this result, we used G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to conduct a power analysis based on the predicted effect size ($d = 0.8$, power = 0.9, alpha = 0.05), and a sample size of 20 was proposed.

2.2. Apparatus

The stimuli were presented on a Dell 19-inch CRT monitor with a spatial resolution of 1024×768 pixels at a 100-Hz refresh rate. The experiment was written in MATLAB, using the Psychophysics Toolbox extensions (Brainard, 1997). Participants sat approximately 60 cm from the monitor during the experiment.

2.3. Stimuli, procedure, and design

Similar to the double rectangles used in the classical cueing paradigm by Egly et al. (1994), the to-be formed objects used in both experiments of this study were double quadrangles, which occupied $3° \times 11°$ of the visual angle (see Fig. 1c, left panel). The two quadrangles were separated by 10°. Each of the quadrangles was divided into two acute angles.

As shown in Fig. 1a, each trial began with a fixation cross presented in the center for 500 ms, followed by a 1000-ms blank screen. Four acute angles were sequentially presented in two frames, each appearing for 200 ms and followed immediately by a mask (presented for 100 ms) and a blank screen (for 1000 ms). Participants were required to memorize the size of all the four

[2] Here the "object" refers to the object formed in VWM, not the object actually presented on the screen.